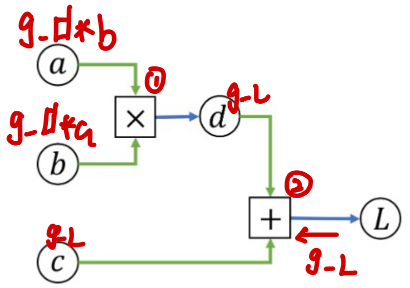


$g\_c = g\_L * \frac{\partial ②}{\partial c} = g\_L$

$g\_d = g\_L * \frac{\partial ②}{\partial d} = g\_L$

$② = c + d$

$\frac{\partial ②}{\partial c} = 1$

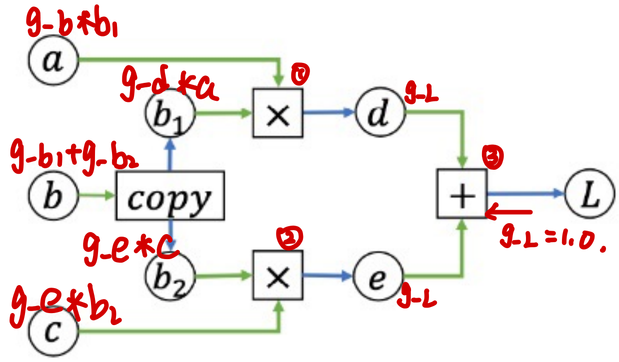
$① = a \times b$

$\frac{\partial ①}{\partial a} = b$

$\frac{\partial ①}{\partial b} = a$

$g\_a = g\_L * \frac{\partial ①}{\partial a} = g\_L * b$

$g\_b = g\_L * \frac{\partial ①}{\partial b} = g\_L * a.$



①: $a \times b_1$

$\frac{\partial ①}{\partial a} = b_1$ $\quad \frac{\partial ①}{\partial b_1} = a$

$g\_a = g\_d * \frac{\partial ①}{\partial a} = g\_b * b_1$

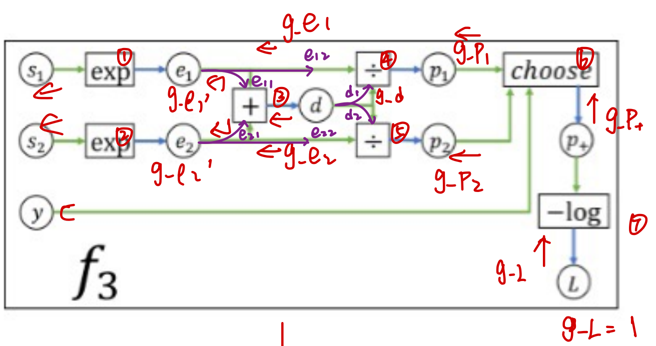
$g\_b_1 = g\_d * \frac{\partial ①}{\partial b_1} = g\_d * a$

②: $b_2 \times c$

$\frac{\partial ②}{\partial b_2} = c$ $\quad \frac{\partial ②}{\partial c} = b_2$

$g\_b_2 = \frac{\partial ②}{\partial b_2} * g\_e = c * g\_e.$

$g\_c = \frac{\partial ②}{\partial c} * g\_e = b_2 * g\_e$



f3 : (required - 10 points)

y is an integer equal to either 1 or 2. You don't need to compute a gradient for y.

The ÷ nodes compute $p_1 = e_1/d$ and $p_2 = e_2/d$.

The choose node outputs outputs p_1 if $y = 1$, and outputs p_2 if $y = 2$.

$⑦ = -\log P_+$

$\frac{\partial ⑦}{\partial P_+} = -\frac{1}{P_+ \cdot \ln 10}$

$g\_P_+ = \frac{\partial ⑦}{\partial P_+} * (g\_L)$

$g\_P_+ = \frac{-g\_L}{P_+ \cdot \ln 10}$

$⑥ = P_1(2-y) - P_2(1-y)$

$= P_1(2-y) + P_2(y-1)$

$g\_P_2 = g\_P_+ * \frac{\partial ⑥}{\partial P_2}$

$= g\_P_+ * (y-1)$

$g\_P_1 = g\_P_+ * \frac{\partial ⑥}{\partial P_1}$

$= g\_P_+ * (2-y)$

$④ = e_1/d$

$g\_e_1 = g\_P_1 * \frac{\partial ④}{\partial e_1}$

$= g\_P_1 * \frac{1}{d}$

$g\_d = g\_P_1 * \frac{\partial ④}{\partial d}$

$= g\_P_1 * e_1 * (-1)(d^{-2})$

$⑤ = e_2/d$

$g\_e_2 = g\_P_2 * \frac{\partial ⑤}{\partial e_2}$

$= g\_P_2 * \frac{1}{d}$

$g\_d = g\_P_2 * \frac{\partial ⑤}{\partial e_2}$

$= g\_P_2 * e_2 * (-1) * (d^{-2})$

$\boxed{g\_d} = g\_P_1 * e_1 + g\_P_2 * e_2$

$③ = e_1 + e_2$

$g\_e_1' = \frac{\partial ③}{\partial e_1} * g\_d$

$= g\_d$

$g\_e_2' = \frac{\partial ③}{\partial e_2} * g\_d = g\_d.$

$g\_e_1 = \quad +$

$g\_e_2 = \quad +$

$① = e^{s_1}$

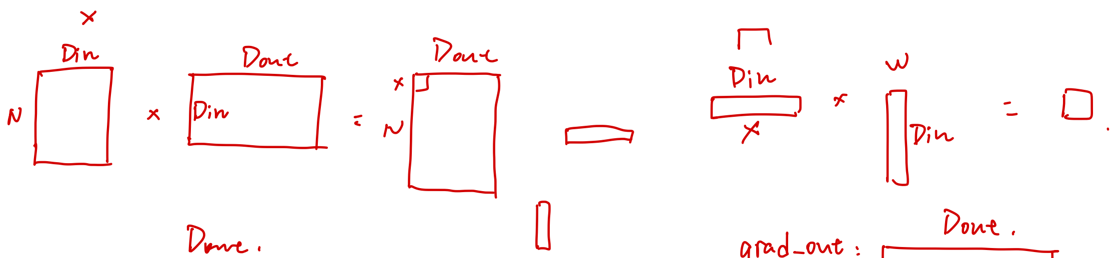
$g\_s_1 = \frac{\partial ①}{\partial s_1} * g\_e_1 = e^{s_1} * g\_e_1$

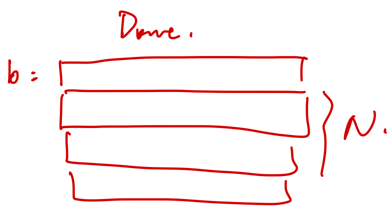
$② = e^{S_2}$

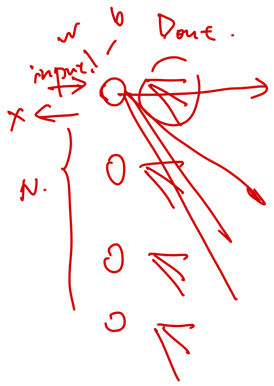
$g\_S_2 = \frac{\partial ②}{\partial S_2} * g\_e_2 = e^{S_2} * g\_e_2$

$g\_y = g\_P_+ * \frac{\partial ⑥}{\partial y}$

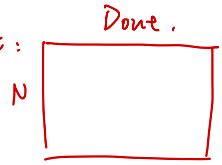
$= g\_P_+ * (-P_1 + P_2)$



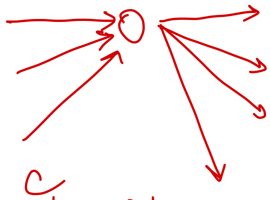


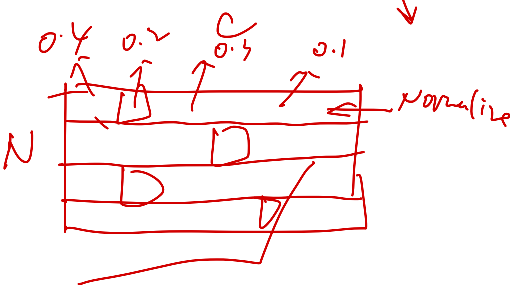


grad_out :



$grad\_x : Dout * \frac{\partial ⑥}{\partial x}$



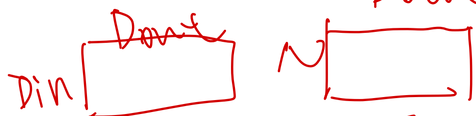


output = X · W + b.

$\frac{\partial output}{\partial w} = x$.

$grad\_w = grad\_out * x$

Dout.

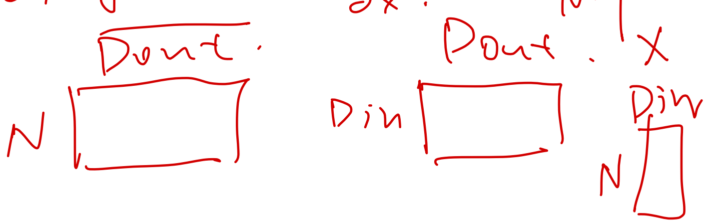


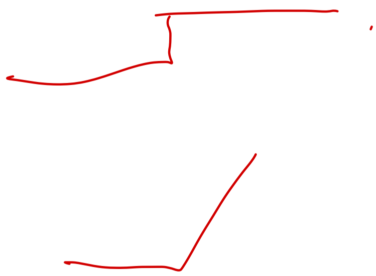
$\frac{\partial output}{\partial x} = w$

$grad\_x = grad\_out * \frac{\partial output}{\partial x}$.

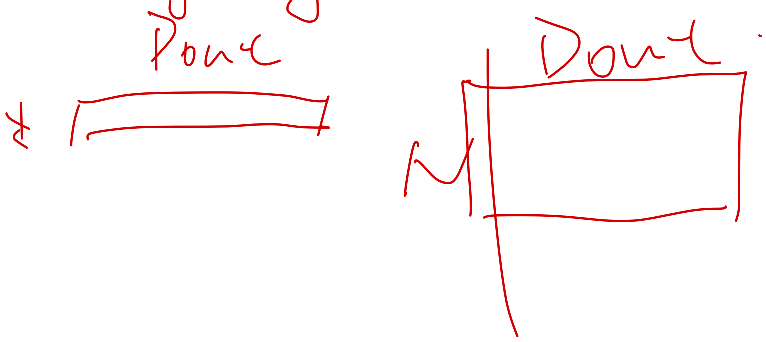
$N \times Din^T$

$N * Dout$





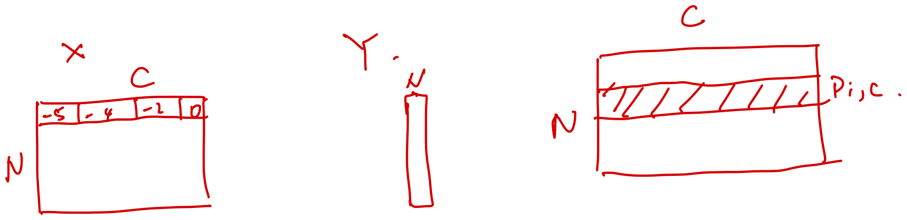
$grad\_y = grad\_out * 1$



$g\_out = g\_in \times f$

$y = x * f$.

$g\_x = grad\_out \times f$



$$\frac{e^{x_{i,c}}}{\sum_{j=1}^{C} e^{x_j}}$$

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log(P)$$

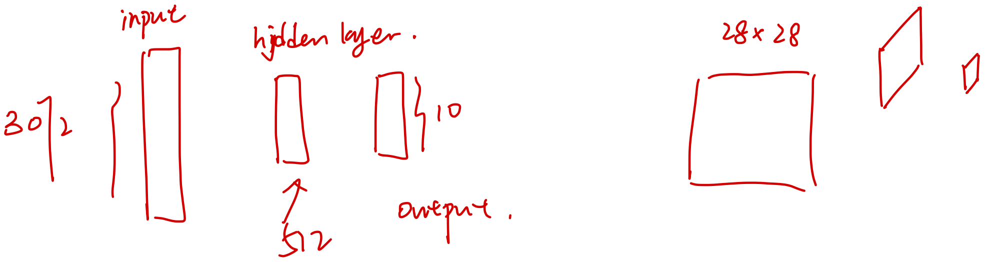
$$\frac{\partial L}{\partial x} = -\frac{1}{n}\sum_{i=1}^{N}\frac{1}{P}\cdot\frac{\partial P}{\partial x}$$

$j=c$.

$$l(w) = \frac{\lambda}{2}\sum_{i} w_i^2$$

$$\frac{\partial l(w)}{\partial w_i} = \frac{\lambda}{2}\cancel{\sum_i}\cdot 2\cdot w_{\cancel{i}}$$

$$= \lambda \cancel{\sum}$$



$1\times28\times28 \xrightarrow{Cov1} 16\times28\times28 \xrightarrow{relu} 16\times28\times28 \xrightarrow{Pool} 16\times14\times14$

$16\times14\times14 \xrightarrow{Cov2} 32\times14\times14 \xrightarrow{relu} 32\times14\times14 \xrightarrow{Pool} 32\times7\times7$

$32\times7\times7 \xrightarrow{fc1} 128 \xrightarrow{fc2} 10.$

$3*32*32 \xrightarrow{Cov1} 16\times32\times32 \xrightarrow{relu} 16\times32\times32 \xrightarrow{Pool} 16\times16\times16$

$16\times16\times16 \xrightarrow{Cov2} 32\times16\times16 \xrightarrow{relu} 32\times16\times16 \xrightarrow{Pool} 32\times8\times8$